Tracking evaluation in discourse

Matthieu Vernier, Stéphane Ferrari

Tracking evaluation in discourse

Matthieu Vernier
GREYC - CNRS UMR 6072
University of Caen, France
mvernier@etu.info.unicaen.fr

Stéphane Ferrari
GREYC - CNRS UMR 6072
University of Caen, France
Stephane.Ferrari@info.unicaen.fr

Abstract

This paper presents a study in the scope of discourse analysis, focussing on how to track evaluation and opinion in discourse. Our aim is to propose a method and a model for automatic analysis of students or teachers opinion in their textual production. A corpus of texts of opinion, collected from different domains, revealed recurrent patterns in the way the authors enunciate their evaluation : lexical, syntactic as well as semantic regularities. Figurative language is also a mean to reinforce the expression of opinion. We propose a NLP model exploiting the previous observations for detection and semantic analysis of the parts of discourse in which authors concentrate their evaluation. This model has been implemented and evaluated during the DEFT07 challenge. We comment the results and discuss how to adapt these model and tool to the E-learning domain.

1 Introduction

Our work adresses the problem of opinion tracking. Opinion is here considered as a sort of expression of evaluation. A first study has been realised on a corpus of books criticals in order to observe the main language and discourse regularities related to evaluation. In section 2, we present the main points of this preliminary work, revealing lexical, syntactic and semantic regularities as well as the importance of rhetorics in the expression of opinion.

In accordance with these observations, we propose a NLP (Natural Language Processing) model and its implementation, in section 3. It is based on a generic structured representation taking into account the subjective part of perception, mostly describing what discourse reveals about what can be called a "social process of culture or knowledge communication". In this representation, the object of the communication, its creator, its addressee and the author of the evaluation are clearly differenciated, in order to identify which part of the social process is the precise object of the evaluation: the creator, the creation process, the message itself, the reception process or the addressee.

In a last section, 4, we present and comment the results of the DEFT07 ("Défi Fouille de Textes") challenge. This year, it proposed an evaluation of opinion tracking systems. Our approach had to be adapted in order to fit the goals. Rather than locally detect and interpretate parts of discourse reflecting evaluation, we had to rate texts one by one, with a score reflecting a global opinion: negative, positive or neutral. We briefly show how we adapted our first NLP system and combined it with a classification tool in this purpose. Then, to conclude, we discuss further possible adaptations for an E-learning application.

2 Previous works

Our contribution is motivated by two different kinds of previous works: a specific study of how evaluation is expressed in discourse, and multiple works on figurative language and rhetorics. In this section, we briefly present these two orientations.

2.1 A first study of evaluation in discourse

Evaluation has recently been studied in the field of corpus linguistics. For instance, in (Hunston and Thompson, 2000), different semantic and grammatical corpus-driven approaches of evaluation are proposed. In (Martin and White, 2005), the authors study the notion of appraisal in a systemic-functional point of view, considering in some way that evaluation is a fundamental motivation of language. In order to be able to propose a NLP model of evaluation, we realized a first study of this phenomenon on a collection of books criticals (Legallois and Ferrari, 2006).

A corpus of 443 comments of customers on books has been collected from the web sites *amazon.fr* and *fnac.com*, containing around 50 000 words. The Internet users are free to post a comment about a book they have read. They also rate it with a system of points, represented by 0 to 5 stars on *amazon.fr* or by an integer between 0 and 10 on *fnac.com*. Different levels of regularities can be observed in the way the authors express their evaluation. We propose the following classification, based on three main classes of regularities. They are not totally without connection one to each other, but they will help separating steps in our methodology and in our model for automatic processing:

- formal regularities, mostly lexical and syntactic ones, some being idiomatic expressions and other ones just reflecting the lexicon of the domain of evaluation ;
- semantic regularities, directly linked to what can be called "experiential frames", the actual context of perception of the evaluated object by the evaluator ;
- enunciative regularities, mostly reflecting how the previous ones act at the discourse level.

In the folloling, we show some examples illustrating these three classes.

2.1.1 Lexical and syntactic regularities

The use of tools for collocations (*Lexico 3* & *Collocates*) revealed formal regularities and a characteristic way of speaking directly related to evaluation. For instance, many sentences start with:
Ce livre est (this book is) + *[evaluation]*

Such phenomenon can be compared to what Hoey calls "colligation" (2005), to characterize a preferred grammatical behaviour, as "collocation" is a preferred lexical behaviour. Simple collocations can be observed, such as:
conseiller vivement (to strongly recommend), *à lire absolument* (to read without condition), *agréable / facile / difficile / rapide à lire* (pleasant / easy / uneasy / quick to read), *s'attendre à mieux* (expecting better), *digne de ce nom* (praiseworthy), *(très) bien écrit* ((very) well written), *à ne pas manquer* (not to miss), *bonne surprise* (good surprise), *sans surprise* (without surprise).

However, idiomatic expressions or verbal phrases are also very common:
passer son chemin (be on your way), *rester sur sa faim* (to be left hungry), *valoir la peine / le détour mériter le détour* (to be worthwhile), *ne pas pouvoir lâcher* (to hang on), *tenir en haleine* (to hold one's breath), *tenir la route* (to hold the road), *être au rendez-vous* (to be at the meeting place)

Without elaborating on these examples, we can see that in a NLP approach, these first observations may help building resources directly expressing evaluation. But it is also interesting to notice that, when observing the variations, for instance around a central verb or a central expression, preferred semantic domains appear to be used. They constitute the second class of regularities.

2.1.2 Semantic regularities

The formal regularities reveal semantic domains regularly used for expressing evaluation. Such domains reflect the values or properties an author evaluates. For instance, affect is one of them, directly linked to how emotion plays a part in evaluation:
on pleure un peu, on rit, on s'émeut!... (we cry a bit, we laugh, we are touched !...)
J'ai pleuré au cours de la lecture de ce livre (I have cried while reading this book)
Ce livre est émouvant (This book arouses emotion)

Many semantic domains also reflect a specific aspect of the reading process. For instance, the notion of "hold of the object over the reader":
on se laisse emporter, on se laisse transporter (we let ourselves being taken away, carried away)
Un livre qui vous vampirise (a book that "sucks the lifeblood out of you")

There is no room in this paper for an exhaustive enumeration of the observed domains. Two main points are important for our study. The first one is that a semantic regularity helps improving the resources for an automatic analysis: it allows us to look for new terms or expressions, even if they were not recurrent in the corpus. The second point concerns rhetorics and figurative language: both the lexical and the semantic regularities obviously show that figurative language has a central role in expressing evaluation.

For instance, the notion of mental grasp is expressed in different ways. The author can give a direct description of his reading process reality: *to feel unable to stop reading, to drop the book* ; or he can use a metaphorical expression, such as one based on the image of "reading is a journey": *this book carried me away*. We consider the use of metaphors and the strength of the images as important clues for an automatic analysis. This point will be discussed further in 2.2.

2.1.3 Enunciative regularities

The last class of regularities we observed consists in strategies of enunciation offered to the author to express evaluation at the discourse level. In this category, we consider for instance expressions characterising the author's commitment: *À mon goût, mon avis, selon moi* (to my opinion, according to me) ; the ones characterising the addressee of the opinion: *vous, on, les fans, tous ceux qui...* (you, one, the fans, the ones who...) ; but also marks of concession, intensity, *etc.*: *certes intéressant au début, mais...* (doubtlessly interesting at the beginning, but...), *Vraiment, véritablement, absolument...* (really, truthfully, absolutely...).

This class of regularities may not be directly related to evaluation, but evaluation can not happen in discourse without this enunciative dimension. In a NLP point of view, we propose to use it as a set of clues for reinforcement of the other clues, in order to determine the intensity of the opinion expressed and to specify to whom this opinion is addressed.

2.2 Figurative language and rhetorics

As previously shown through a few examples, figurative language is central in the expression of evaluation. Many metaphors are used to describe the way the subject lives its reading: "reading is a journey", "the book is a living creature able to charm, to put a spell or to hold the reader", "the book's content is food, to eat, to drink or to digest". Such metaphors can be compared to the notion of conceptual metaphor introduced in (Lakoff and Johnson, 1980). Different propositions have already been made to integrate this notion in NLP systems: a model for lexical resources and their exploitation were presented in (Martin, 1991), recent works also integrate the notion in WordNet-like resources, as in (Alonge and Castelli, 2003).

Other works on metaphors and analogy proposed different structured representations for automatic processing, mostly based on analogy, as in (Gentner, 1983; Falkenhainer et al., 1989) or (Fass, 1997). In our laboratory, a project focussed on the semantics of conceptual metaphors, leading to tools for detection and interpretation, as presented in (Beust et al., 2003; Perlerin et al., 2003; Perlerin et al., 2005). All these models have their own purposes. In our study, we already know most of the images or conceptual metaphors used. Therefore, it is not necessary to find back metaphorical meanings through complex computing of analogies. But all the works based on conceptual metaphors are relevant, depending on the precision of the interpretation we are looking for.

A last word must be said about figures of speech and rhetorics in general. Metaphor is not the only figure used in the studied corpus. It is probably the easiest one to observe through the semantic regularities, but we also noticed numerous emphasis, overemphasis, as well as understatements, which are all closely related to the expression of evaluation, adding strength to the opinion. This kind of figures can also add important information in a NLP system dedicated to evaluation. In (Klinkenberg, 2001), the author explains how rhetorics is in some way inherent to discourse, whatever the kind of text. In (Ferrari, 2006), the author presents the most common figures influencing the semantics of discourse. For the purpose in this specific study, taking them into account will help balancing the strength of the expressed opinion.

All the regularities or phenomenons detailed in this section are taken into account in the model and its implementation presented next. This model was

previously built for opinion tracking in critical of cultural objects (books, movies, comics, music...). We will show in a last section how it can be adapted to the E-learning domain, in order to help interpreting evaluation given e.g. by teachers on students works or by students on courses and exercises.

3 NLP Model and Implementation

In this section, we first propose a NLP model, based on a generic structured representation of the "social process of culture or knowledge communication", taking into account the subjective part of perception discourse reveals about an author. This representation distinguishes between the creator, the object of the communication (cultural one or knowledge) and the addressee, in order to spot which element is the precise object of an evaluation.

Next, we present an implementation of the NLP model. It is the one which has been used in our participation to the DEFT07 challenge. It shows how previous works and a previous version of this model can be applied to different contexts of culture or knowledge communication.

3.1 A social process representation

Interpreting a text requires to consider knowledge about the reality this text talks about. In (Ricœur, 1975), the author explains how this particularly applies to figurative language and metaphors. We here adopt the same point of view in order to organise the interpretation of evaluation in the specific context of culture or knowledge communication. We propose an abstract representation of this social process, shown in figure 1. It was initially built for describing cultural contexts, but we argue it can also apply to a global communicative context. In particular, when considering the E-learning domain, the creator can be for instance the teacher, the object a course, and the addressee a student.

We consider that discourse does not describe a reality in itself but rather how this reality is perceived by the enunciator. When transmiting opinion about a thing, we do not always formulate our idea warning that this is only our own feeling. Like a child who does not appreciate a food and say "It's not good" instead of "i don't like this food", we sometimes describe the reality as we feel it. In discourse, the enunciation strategies may contain clues about the author's feelings and subjectivity. His vision of reality can be coloured or oriented in different ways, reflecting a positive or a negative perception. This can be observed in the DEFT07 corpus (see 4 for details).

Examples :
(1) *"Coppola choisit de donner sa chance au jeune Robert de Niro qui* ***compose*** *un Vito Corleone la présence* ***magnétique*** *et* ***envoûtante****."*("Coppola chooses to give his chance to the young Robert de Niro who **composes** a Vito Corleone with a **magnetic attraction** and a **captivating** presence".)
(2) *"On peut donc se laisser tenter par ce Bombon,* ***sucrerie douce-amère*** *qui se laisse* ***déguster*** *avec un plaisir non dissimulé ".* ("Let's try this Bombon, **bitter sweet sugary** to **taste** with unconcealed pleasure".)

The way of catching reality by the author of (1) shows that he perceives the relation between Robert de Niro and an object (here, his character) as analogous to the relation between a composer and his artistic piece. The verb *to compose* links a creator (the composer) and an object (his musical piece) which can be the support for a message (see figure 1). This metaphor subjectively colours the actor as a musical artist and provides by this mean a positive lightning to the cultural object creation process. Moreover, the terms *magnetic attraction* and *captivating* are used to describe a sensible experiment between the message carried by the cultural object itself and one or several potential addressees. They a "mental grasp" between the *addressee(s)* and the *object*.

In (2), expression *bitter sweet sugary* is a nominal anaphora which, according to (Schnedecker, 1997), reveals a subjective judgement at the discourse level. *Sugary* refers to an object (here, the movie) but through a sensory perception image. *Bitter* and *sweet* reinforce the fact that the author is talking about the movie, but especially the relation with *addressee(s)*. The sense of taste is often used to formulate our positive or negative relation with an object when we are in a reception position.

Considering the elements and relations involved in the social process is necessary to determine how the reality is perceived by the enunciator. The ab-

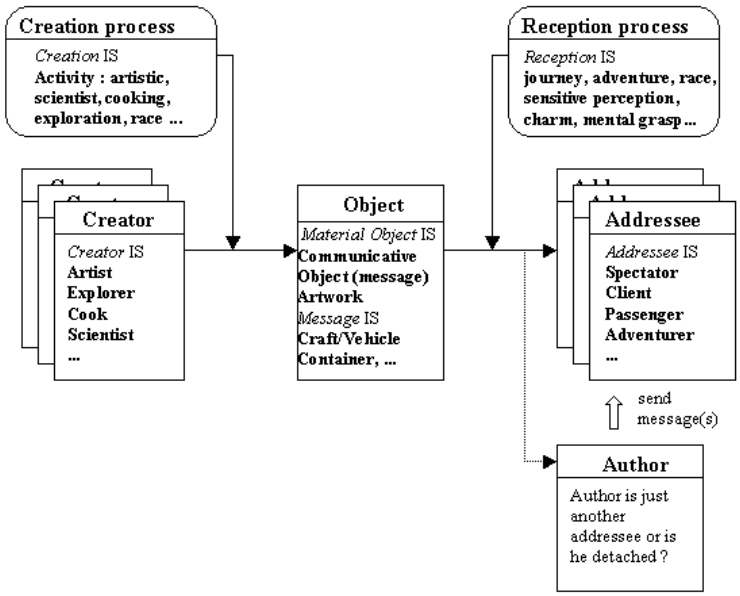

Figure 1: Social Process Modelisation
Author: “Critic”, “Teacher”, “Student”...
Creator: “Movie maker”, “Actor”, “Fiction writer”, “Comic author”, “Artist”, “Teacher”, “Student”...
Object/Message: “Movie”, “Book”, “Comic”, “Spectacle”, “Homework”, “Course”...

stract representation given in figure 1 aims to specify how linguistic phenomenons like metaphors or metonymies enable to recover the subjective point of view of the enunciator on these objects and relations. In our NLP resources, for a specific conceptual metaphor, figurative terms are linked to the corresponding part of reality, and the positive or negative connotations are encoded as well.

Lastly, detecting an opinion about an element of the structure does not mean the same opinion applies to the other elements, and it does not necessarily reflect the global opinion of the author. An author can focus his argumentation on one element or one relation only. A critic can compliment how a filmmaker *builds* his movie, but he can say, in the same discourse, that it *leads nowhere* and that the message does not reach the spectator. Such a metaphor would be a positive point for the creation process but negative for the relation *Object/Addressee*. Most of the times, authors organise different parts in their discourse, giving different opinions on different aspects. It is interesting to take the position of these parts of text into consideration. For instance, our previous example (1) appears at the very end of a critic. Concluding by this sentence, the author points that the imperfection in the building of the film is less important than the spectator satisfaction. Therefore, our NLP model also takes positions of textual elements into account to reinforce or moderate the expressed opinion. Most of the times, introduction and conclusion are, for instance, considered as more important than other positions.

3.2 Implementation in LinguaStream

We use the *LinguaStream* platform, developed at the GREYC, (Ferrari et al., 2005; Widlöcher and Bilhaut, 2006). It provides an integrated development environment for designing complex NLP systems. It relies extensively on XML, and demonstrates a practical use of this standard and surrounding ones for NLP, taking particular attention on semantic concerns. Thus, we developped a NLP system (see figure 2) which can be automatically applied to each text of a corpus. It annotates parts of text related to evaluation or subjectivity with semantic features. The system is mostly based on lexical resources and a Prolog grammar for detecting lexical and syntactic clues. It also uses preliminary NLP modules for tokenizing and tagging the texts, which are already integrated in the platform.

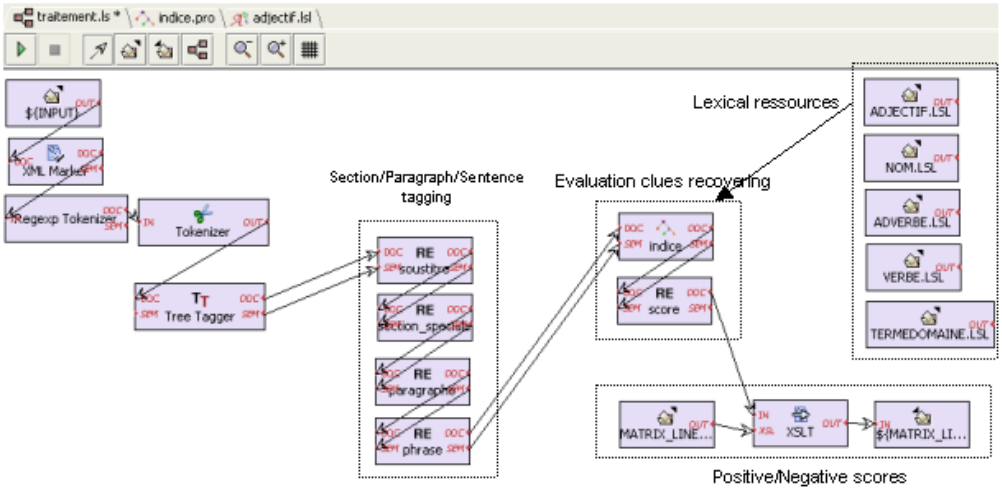

Figure 2: Implementation in Linguastream

Examples of detected clues :
[evaluation] *and [modifiers]*
"Cette BD est une [véritable] [grande] ***[réussite]****."*
("This comic book is a real big success")
"Un roman [très] ***[ennuyeux]****."*
("A very boring novel")
"Ce film nous ***[plonge]*** *dans une atmosphère sombre."*
("This film dips us into a dark atmosphere")
"L'auteur ***[fait danser]*** *les mots."*
("The author makes the words dance")

Attributes are affected to each lexical entry, corresponding to the semantics features relevant for our purpose (positive or negative aspect, conceptual metaphors involved...).
Examples of lexical resources :
"réussite" (success) {clue: {evaluation: 1}}
"ennuyeux" (boring) {clue: {evaluation: -1, evaluated: reception}}
"plonger" (to dip) {clue: {evaluation: 1, metaphor: mental-grasp, evaluated: reception}}
"très" (very) {clue: {modifier: intensity}}

A Prolog grammar is used to give a weight to each clue found in a text, computing a score according to the presence of enunciative reinforcements, adverbs of intensity, concession... (see (Vernier et al., 2007) for details). This part of the NLP system is the main module which can be also applied to the E-learning domain, allowing to detect local evaluation applying on specific parts of our social process representation.

At the end of this processing, a last component computes global scores for the whole text, depending on the position of each clue (e.g. balancing the ones found in introduction or in conclusion). Finally, a classifying automatic process based on extraction rules determines if a text is positive, negative or neutral. These last steps are not part of our model but just the way to adapt it to the DEFT07 challenge, using the same approach as in (Widlöcher et al., 2006): the results of the NLP system are used as entries for an automatic classifier. Next section presents the results of this challenge.

4 Results and Perspectives for eLearning

Through the DEFT07 Challenge, our system has been validated on different corpora (see (Vernier et al., 2007) for further discussion). Initially designed for cultural objects criticals (movies, books, comics, play, etc), a simplification of our current model has also been applied to a corpus of scientific papers reviews and to an other corpus of video games evaluations. In the first stage of the challenge, corpora in each domain were given to the participants in order to train their systems (NLP resources or rules automatic acquisition...). In the second and last stage, the systems were evaluated on other corpora constituted the same way as the practice ones. The results show some interesting regularities.

Number	Corpus type	Score
1	Cultural Objects Criticals	0,457
2	Video Games Evaluations	0,506
3	Scientific Papers Reviews	0,474

Figure 3: Results for DEFT07 Challenge

The results seem almost independant of the corpus type. Best results are obtained on homogeneous corpora (2 and 3). The lexicon of these corpora were easier to study in the amount of time left for the challenge. It appears that corpus 1 was the most heterogeneous, including criticals of books, comics, records, films. The images used were too numerous and the lexical or syntactic regularities not as recurrent as in a homogeneous corpus to provide stable resources to our system. Further linguistic study of the material would be necessary to increase the resources, and, as a consequence, the number and the quality of clues found in this type of corpus.

It seems that some rhetorical figures are common to a communicative process, in culture or knowledge transmission, and can be directly applied to the particular context of E-learning. In an E-learning purpose, we think that most of rhetorical figures draw equivalent relations in a similar model. In a certain way, a *Teacher* can be seen as a *Creator* of an *Object*: his *Course*. In that particular case, the *Student* will be the *Author/Addressee* and will comment or evaluate his relation with the *Course*. He may use different conceptual metaphors to express his evaluation, like in :
(3) *"Ce concept est **flou**."*("This concept is **blurry**."
(4) *"Je suis **perdu** dans l'exercice."*("I **got lost** in this exercise.")

Proposition (3) uses lexicon from sensory perception domain, and especially sight. We can suppose that a metaphor like "You can only see what you understand" is paramount in this context. In the same idea, an analysis of *light* lexical field could be interesting. Terms like *clair* ("clear"), *lumineux* ("bright"), *obscur* ("unlit"), *flou* ("blurry") are often used to evaluate how we understand something. Conceptual metaphors "Bright IS Good" and "Dark IS Bad" are two axis to take into consideration to retrieve the positive or negative aspect.

Proposition (4) is an example of metaphor involving a path or an adventure. This kind of image should help to know that the author is speaking of the reception relation (here, the object is the exercise) (figure 1). The fact of being *lost* means a negative evaluation to this relation.

In the same manner, the *Student* can become the *Creator* if he makes an exercice or a homework, and the *Teacher* the *Author/Addressee* and would give his opinion by evaluating the student's work.
Examples :
(5) *"Tu n'as pas assez exploré le cours."*("You didn't explore the course enough.")

The different roles in our social process representation can easily be transferred to the context of knowledge communication. The methodology we followed can also be applied to specify resources relevant in the E-learning domain: a corpus study should reveal images and formal regularities as well as it did in other domains. For all these reasons, we think opinion and subjectivity in E-learning could probably be detected in the textual production of teachers and students as well as in cultural objects criticals.

5 Conclusion

We presented a model for opinion tracking, mostly based on textual clues and discourse analysis. A first study of an observatory corpus revealed recurrent linguistic patterns, both lexical, syntactical and semantic regularities, as well as the use of multiple figures, mostly metaphors, to express evaluation. We thus proposed a NLP model integrating these regularities, in relation to a representation of the social process underlying the specific domain of cultural communication. The whole led to an implementation that has been evaluated during the DEFT07 challenge. The results on different corpora, not always opinions on cultural objects, tends to show a good adaptability of the whole model.

In the scope of E-learning, it seems possible to apply a similar representation of a social process of knowledge communication, where the teacher and the student can take the place of the author and the addressee. In this perspective, the current NLP system we developped can be easily adapted, by modification of the resources, taking into account the

specificities of the language used in the E-learning context. Such adaptation would require a preliminary study on a corpus, in order to determine what images are used in this domain, as well as what formal regularities.

References

A. Alonge and M. Castelli. 2003. Encoding information on metaphoric expressions in wordnet-like resources. In J. Barnden, S. Glasbey, M. Lee, and A. Wallington, editors, *Proceeding of the ACL 2003 Workshop on the Lexicon and Figurative Language*, pages 10–17.

P. Beust, S. Ferrari, and V. Perlerin. 2003. NLP model and tools for detecting and interpreting metaphors in domain-specific corpora. In D. Archer, P. Rayson, A. Wilson, and T. McEnery, editors, *Proceedings of the Corpus Linguistics 2003 Conference*, pages 114–123, UCREL, Lancaster University, U.K.

B. Falkenhainer, K.D. Forbus, and D. Gentner. 1989. The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41(1):1–63. D.G. Bobrow et M. Brady (eds), Elsevier Science Publishers B.V., Amsterdam, Netherlands.

D. Fass. 1997. *Processing Metonymy and Metaphor*. Ablex Publishing Corporation, Greenwich, Connecticut, U.S.A.

S. Ferrari, F. Bilhaut, A. Widlcher, and M. Laignelet. 2005. Une plate-forme logicielle et une démarche pour la validation de ressources linguistiques sur corpus : application l'évaluation de la détection automatique de cadres temporels. In G. Williams, editor, *Actes des 4èmes Journées de la Linguistique de Corpus*. 15-17 septembre 2005, Lorient, France, Université de Bretagne-Sud. To be published.

S. Ferrari. 2006. Rhétorique et compréhension. In Gérard Sabah, editor, *Compréhension des langues et interaction*, Traité IC2, série Cognition et traitement de l'information, chapter 6, pages 195–224. Hermès Sciences, Lavoisier, Paris.

D. Gentner. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7:155–170.

M. Hoey. 2005. *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge.

S. Hunston and G. Thompson, editors. 2000. *Evaluation in Text. Authorial Stance and the Construction of Discourse*. Oxford University Press.

J.-M. Klinkenberg. 2001. L'argumentation dans la figure. *les Cahiers de praxématique*, 35:17–38.

G. Lakoff and M. Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago, U.S.A.

D. Legallois and S. Ferrari. 2006. Vers une grammaire de l'valuation des objets culturels. In *Actes d'ISDD06, colloque international Discours et Document, Schedæ, 2006*, pages 57–68. Presses universitaires de Caen, France.

J. Martin and P. White. 2005. *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan Hardcover.

J. Martin. 1991. MetaBank : a Knowledge-Base of Metaphoric Language Conventions. In D. Fass, E. Hinkelman, and J. Martin, editors, *Proceedings of the IJCAI Workshop on Computational Approaches to Non-Literal Language*, pages 74–82. Sydney, Australia.

V. Perlerin, P. Beust, and S. Ferrari. 2003. Computer-assisted interpretation in domain-specific corpora: the case of the metaphor. In *Proceedings of NODALIDA'03, the 14th Nordic Conference on Computational Linguistics*, University of Iceland, Reykjavík, Iceland.

V. Perlerin, S. Ferrari, and P. Beust. 2005. Métaphores et dynamique sémique : expériences sur corpus. In G. Williams, editor, *La linguistiqe de corpus*, pages 323–336. Presses Universitaires de Rennes, France.

P. Ricœur. 1975. *La métaphore vive*. Seuil, Paris, France.

C. Schnedecker. 1997. *Nom propre et chaînes de référence*. Universit de Metz - Klincksieck, Paris, France.

M. Vernier, Y. Mathet, F. Rioult, T. Charnois, S. Ferrari, and D. Legallois. 2007. Classification de textes d'opinions : une approche mixte n-grammes et sémantique. In *Troisième DÉfi de Fouille de Textes (DEFT'07), Plate-forme AFIA (Association Française d'Intelligence Articifielle) 2007*. 2-6 juillet, Grenoble, France.

A. Widlöcher and F. Bilhaut. 2006. LinguaStream: An Integrated Environment for Computational Linguistics Experimentation. In *Proceedings of EACL 2006, the 11th Conference of the European Chapter of the Association of Computational Linguistics (Companion Volume)*, pages 95–98. Trento, Italy, April 15-16 2006.

A. Widlöcher, F. Bilhaut, N. Hernandez, F. Rioult, T. Charnois, S. Ferrari, and P. Enjalbert. 2006. Une approche hybride de la segmentation thématique : collaboration du traitement automatique des langues et de la fouille de texte. In *Deuxième DÉfi de Fouille de Textes (DEFT'06), Semaine du Document Numérique (SDN'2006)*. 18-22 septembre, Fribourg, Suisse.